

Experimentelle Kompetenz messen?!

Nico Schreiber*, Heike Theyßen*, Horst Schecker⁺

* Lehrstuhl für Didaktik der Physik, TU Dortmund,

⁺ Institut für Didaktik der Naturwissenschaften, Abt. Physikdidaktik, Universität Bremen

(Eingegangen: 05.05.2009; Angenommen: 09.09.2009)

Kurzfassung

Experimentelle Kompetenz als Teil des Kompetenzbereichs „Erkenntnisgewinnung“ soll gemäß der Bildungsstandards im Unterricht gefördert werden. Für die Diagnose experimenteller Kompetenz müssen daher auch geeignete Werkzeuge zur Verfügung stehen. Es ist zu bezweifeln, dass schriftliche Wissenstests valide genug sind, um die ganze Breite experimenteller Fähigkeiten, einschließlich der tatsächlichen Durchführung von Experimenten, zu erfassen. Experimentaltests mit Realexperimenten hingegen sind in Durchführung und Auswertung für den großflächigen Einsatz in Lernstandserhebungen oder PISA-Tests sehr aufwändig. Eine Alternative könnten Experimentaltests mit virtuellen Medien, wie Simulationsbaukästen für experimentelle Handlungssituationen, sein. Ob solche Verfahren zur Diagnose geeignet sind, soll durch das hier beschriebene Vorhaben überprüft werden. Den Erhebungen liegt ein Modell experimenteller Kompetenz zugrunde, welches durch Indikatoren operationalisiert ist.

Für die Kompetenztests wurden zwei experimentelle Aufgabenstellungen in je drei Testversionen (schriftlich, Realexperiment und Simulationsbaukasten) entwickelt. In einer Pilotstudie bearbeiten Schüler der 10. bis 12. Klassenstufe die Aufgabenstellungen in jeweils unterschiedlichen Testversionen. Die Vorstudie dient der Erprobung und Optimierung der Übungs- und Begleiterhebungsmaterialien, der Kompetenztests sowie der Auswertungsmethoden. Im Rahmen der Hauptstudie werden die mit den verschiedenen Testversionen erzielten Kompetenzzuweisungen auf Korrelationen hin untersucht. In diesem Artikel werden der theoretische Hintergrund, speziell das zugrunde gelegte Modell experimenteller Kompetenz sowie das Untersuchungsdesign vorgestellt.

1. Motivation und Zielsetzung

Sowohl in den nationalen Bildungsstandards [1] als auch in daran anschließenden fachdidaktischen Kompetenzmodellen (Schecker & Parchmann [2], Neumann et al. [3]) wird experimentelle Kompetenz explizit als eine Komponente naturwissenschaftlicher Kompetenz im Bereich „Erkenntnisse gewinnen“ oder „naturwissenschaftliches Arbeiten“ ausgewiesen. Neben Methoden zur Erhebung von Fachwissen werden also in zunehmendem Maße Methoden zur Erhebung experimenteller Kompetenz benötigt. Dies gilt sowohl für die Kompetenzdiagnostik auf Individualebene durch die einzelne Lehrperson als auch für das Assessment zur Bilanzierung von Lernergebnissen mit Schulleistungsstudien. In Studien wie TIMSS und PISA wurden in Deutschland bisher experimentelle Kompetenzen nicht durch Experimentaltests erhoben, anders als z.B. in der Schweiz mit den TIMSS-Experimentiertests (vgl. [4], [5]) oder im Projekt HarmoS, (vgl. [6]). Ein praktischer Grund ist der logistische Aufwand, die entsprechenden Versuchsmaterialien einheitlich bereitzustellen. Deshalb wird oft versucht, die Experimentierfähigkeiten der Schülerinnen und Schüler durch schriftliche Wissenstests zu messen. Wir gehen davon aus, dass diese Diag-

nostik nicht hinreichend valide ist, weil prozessbezogene Aspekte, die wesentlich für die experimentelle Vorgehensweise sind, nicht erfasst werden. Wünschenswert wäre eine alternative Erhebungsmethode, die eine Erfassung der prozessbezogenen Aspekte zulässt, aber weniger aufwändig in Durchführung und Auswertung ist als Experimentaltests mit Realexperimenten.

Das zentrale Ziel des Projekts „Diagnostik experimenteller Kompetenz“ (eXkomp)¹ ist daher die Entwicklung und Validierung von Testverfahren, die handlungs- und prozessbezogene Aspekte experimenteller Kompetenz erfassen, empirisch abgesichert und gleichzeitig potenziell großflächig einsetzbar sind. Simulationsbaukästen (vgl. z.B. [7]), die vielfältige experimentelle Handlungsmöglichkeiten erlauben, könnten die Grundlage für eine solche Alternative sein.

Voraussetzung für die Entwicklung und Validierung eines solchen Messverfahrens ist ein ausdifferenziertes und operationalisiertes Konstrukt experimenteller Kompetenz. Auf dieser Basis kann überprüft wer-

¹ Das Projekt wird durch die Deutsche Forschungsgemeinschaft (DFG) unterstützt.

den, welche Erhebungsmethoden geeignet sind, um im Rahmen des theoretischen Konstrukts experimentelle Kompetenz zu messen.

2. Modell experimenteller Kompetenz

Experimentieren ist sehr vielfältig und es gibt verschiedene Vorstellungen, was unter dem Begriff des Experimentierens zu verstehen ist [8]. Wir wollen uns hier an der sehr spezifischen Sicht bereits bekannter Kompetenzmodelle aus der (fachdidaktischen) Forschung orientieren und die Situation des typischen selbstständigen Experimentierens, z.B. in der Schule, berücksichtigen.

Eine Grundlage für die Ausdifferenzierung der experimentellen Kompetenz stellt ein von Hammann [9] vorgeschlagenes Modell dar. Das Modell orientiert sich an der Darstellung des naturwissenschaftlichen Erkenntnisweges bei Klahr [10] und baut gleichzeitig auf empirischen Befunden auf. Hammann unterscheidet nach Klahr drei Teilkompetenzen: das Suchen und Aufstellen von Hypothesen, die Planung und Durchführung von Experimenten sowie die Analyse der experimentellen Ergebnisse. Innerhalb dieser Bereiche differenziert er zwischen vier nach der Systematik der Vorgehensweise gestuften Kompetenzniveaus.

Ein ähnliches Modell mit vier Teilkompetenzen schlägt Mayer vor [11]. Es enthält als vierte Teilkompetenz die Formulierung naturwissenschaftlicher Fragestellungen. Die Durchführung von Experimenten wird mit schriftlichen Wissenstests als nicht erfassbar aufgefasst und in diesem Modell nicht als eigene Teilkompetenz berücksichtigt. Möller et al. [12] finden innerhalb der Teilkompetenzen fünf gestufte Kompetenzniveaus.

Im Unterschied zu diesen Arbeiten, die sich insbesondere mit der Hypothesenbildung und Interpretation beschäftigen, legen wir den Fokus auf die *Durchführung* von Experimenten, einschließlich des Prozesses des Aufbaus und Messens als zentralen und fachbezogenen Teil des Experimentierens. Die der Durchführung vorausgehende Planung und die Auswertung der experimentellen Ergebnisse werden als Randbereiche mit erfasst und ergänzen den Dreischritt nach Hammann und Klahr.

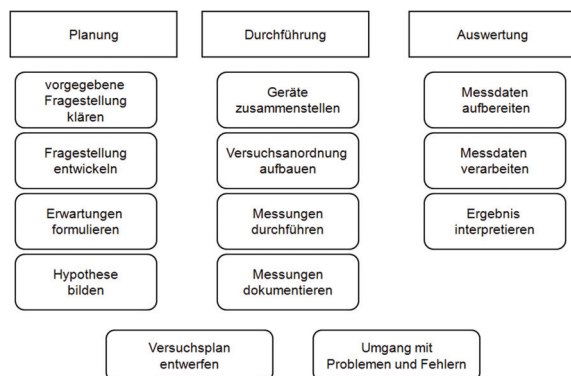


Abb.1: Modell experimenteller Kompetenz

Abbildung 1 zeigt die drei übergeordneten Bereiche „Planung“, „Durchführung“ und „Auswertung“ mit den zugeordneten Merkmalen des Experimentierprozesses. Die drei Bereiche sowie die zwei nicht eindeutig zugeordneten Merkmale an den Schnittstellen werden im Folgenden erläutert.

2.1. Planung

Wenn Forscher ein unbekanntes Phänomen entdecken, werden sie zunächst *Fragestellungen entwickeln*, die eine genauere Klärung bzw. Präparation des eigentlichen Phänomens ermöglichen. Erst nach intensiver Auseinandersetzung und auf Grundlage theoretischen Wissens ist die Formulierung begründeter Annahmen über Wirkungszusammenhänge („*Hypothesen*“) möglich. Bei im Grundsatz bekannten Phänomenen kommt es in der Wissenschaft oftmals gar nicht darauf an, Fragestellungen oder Hypothesen selbst zu entwickeln, sondern Aufträge mit gegebenen Fragestellungen und Hypothesen sorgfältig abzuarbeiten und gegebenenfalls begrenzt weiterzuentwickeln, z.B. hinsichtlich der experimentellen Verfahren. Dies gilt noch für viele Diplomarbeiten oder auch Promotionsvorhaben. Hier muss sich der Durchführende lediglich in die Fragestellungen einarbeiten und das Ziel der Untersuchung klären, um es zu einem eigenen Ziel machen zu können.

In Lehr-Lernsituationen wird der Generierung von Hypothesen und eigenen Fragestellungen oftmals in einer idealisierten Vorstellung von (normal-) wissenschaftlicher Forschung ein zu großer Stellenwert eingeräumt. In unserem Modell, das stärker als die Ansätze von Klahr, Hammann oder Mayer auf den Prozess des Experimentierens ausgerichtet ist, wird daher der Klärung vorgegebener Fragestellungen durch den Schüler eine mindestens ebenso große Bedeutung zugemessen wie der Entwicklung eigener Untersuchungsziele oder sogar Hypothesen. Erwartungen über den Ausgang eines Experiments sind allerdings notwendig, um „Fehler“ erkennen zu können (s.u.). Erwartungen sind jedoch niederschwelliger als theoretisch begründete Vorhersagen des Versuchsausgangs.

Zwischen der Planung und der Durchführung einer experimentellen Untersuchung steht im Modell zunächst der gedankliche und dann der reale Entwurf einer Versuchsanordnung und der Entwurf eines Versuchsplans. Sie sind als wichtige Schnittstelle beiden übergeordneten Bereichen des klassischen naturwissenschaftlichen Erkenntnisweges zuzuordnen. Ein wesentlicher Teilaspekt der Entwicklung eines Versuchsplans ist die Parameterkontrolle. Diese kann (und sollte) sowohl explizit in der Planung berücksichtigt werden als auch implizit in der Durchführung.

2.2. Durchführung

Wenn ein physikalisches Experiment durchgeführt werden soll, müssen – auf Basis des Umsetzungsplans – geeignete Geräte zusammengestellt sowie die Versuchsanordnung aufgebaut und getestet werden. Der Entwurf des Versuchsplans und die Schritte bis zu einem funktionstüchtigen Experiment bilden keine lineare Abfolge. Es handelt sich in der Regel um spiralförmig durchlaufene (Teil-) Prozesse mit gegenseitigen Modifikationen. Eindeutige Expertenslösungen für das Durchlaufen der Schleifen lassen sich schwer angeben. Schließlich werden die Messungen durchgeführt und dokumentiert.

2.3. Auswertung

Wenn die Messungen abgeschlossen sind, müssen die Messdaten gegebenenfalls tabellarisch oder grafisch aufbereitet werden; es müssen Zusammenhänge zwischen physikalischen Größen untersucht (Verarbeitung der Messdaten) und die Ergebnisse interpretiert, d.h. auf die Fragestellung, Erwartung oder Hypothese bezogen werden.

Der Umgang mit Fehlern ist wieder eine Schnittstelle und sowohl der Durchführung als auch der Auswertung zuzuordnen. Hierzu gehört der Umgang mit nicht funktionierenden Aufbauten bei der Durchführung ebenso wie das Erkennen von fehlerhaften Messwerten, Interpretationen oder Ansätzen, die nicht zur Beantwortung der Fragestellung geeignet sind.

Um es noch einmal zu betonen: Die obige Darstellung soll nicht den Eindruck vermitteln, es gäbe eine strenge Reihenfolge innerhalb des Dreischritts. Das Modell geht vielmehr davon aus, dass alle Merkmale in einem Experimentierprozess in verschiedenen Reihenfolgen, gegebenenfalls iterativ und auch nicht immer vollständig auftreten können. Das gilt insbesondere für Maßnahmen im Umgang mit Problemen und Fehlern und Variationen des Versuchsplans (vgl. Abb. 1). Gegenstand des laufenden Forschungsprojektes ist es u.a. charakteristische Abläufe zu identifizieren und gegebenenfalls mit dem Erfolg des Experimentierens zu korrelieren. Erste Ergebnisse (vgl. [13], [14]) deuten jedoch darauf hin, dass der Verlauf des Experimentierens kaum mit dem Erfolg zusammenhängt.

3. Testformate

3.1. Bislang übliche Testformate

Tests zur Erhebung experimenteller Kompetenz in den Naturwissenschaften liegen vor, greifen jedoch bislang fast ausschließlich auf zwei Verfahren zurück: Experimentaltests mit Realexperimenten und schriftliche Wissenstests. Bei den Experimentaltests ist zu unterscheiden, ob die Auswertung *prozessbezogen*, d.h. durch Handlungsanalysen, oder *produktbezogen*, z.B. anhand schriftlicher Arbeitsergebnisse, erfolgt. Experimentaltests mit prozessbezogener Auswertung liegen von Walpuski [15] und Neumann

[16] vor. Walpuski erstellt aus Videoanalysen Prozessgrafiken. Er orientiert sich wesentlich an einem Dreischritt (Idee, Experiment, Schlussfolgerung), der Parallelen zu Hammanns Strukturierung experimenteller Kompetenz enthält. Neumann filmt und analysiert kategoriegeleitet die Experimentaltests. Hierbei verwendet Neumann ein Kategoriensystem nach C. von Aufschnaiter [17], das zwischen explorativer, intuitiv regelbasierter und explizit regelbasierter Vorgehensweise unterscheidet. Diese Differenzierung für den Bereich der experimentellen Durchführung weist Parallelen zu der bei Hammann [9] vorgenommenen Unterscheidung zwischen systematischem und unsystematischem Umgang mit Variablen auf.

In den TIMSS-Experimentiertests [18] wurden einfache experimentelle Aufgaben von den Probanden bearbeitet (z. B. Kraft-Dehnungs-Zusammenhang bei einem Gummiband). Es wurden nur die Ergebnisse („Produkte“) ausgewertet. Ergänzend zu schriftlichen Wissenstests werden die Experimentaltests bei „HarmoS“ [6] ebenfalls produktbezogen ausgewertet.

Mit Ausnahme von TIMSS und HarmoS erfolgt die Untersuchung großer Probandengruppen bisher in der Regel durch schriftliche Wissenstests. Zu diesen werden hier auch solche Varianten gezählt, bei denen Experimente nicht schriftlich beschrieben, sondern vorgeführt werden und anschließend Fragen zu Vorgehensweise und Interpretation zu beantworten sind ([19]; vergleichbar auch mit Experimentalaufgaben in Abiturklausuren). Zur Erfassung naturwissenschaftlicher Arbeitsweisen mit schriftlichen Wissenstests wurde von Henke und Sumfleth [20] ein Testinstrument entwickelt und in zahlreichen Untersuchungen erprobt und weiterentwickelt (z.B. [21], [22]). Zur Erhöhung der Auswertungsobjektivität wurde der Test als Multiple-Choice-Test konzipiert. Um primär Kenntnisse über naturwissenschaftliche Arbeitsweisen zu erfassen, wird das Fachwissen ähnlich wie bei den PISA-Aufgaben [23] in der Aufgabe vorgegeben. Auch die bei „HarmoS“ eingesetzten schriftlichen Wissenstests (vgl. [6], S. 106) sind nach PISA – Vorbild konstruiert.

3.2. Ein alternatives Testformat

Ein alternatives Testformat für prozessbezogene Kompetenzen kann auf einem Simulationsbaukasten basieren. Abbildung 2 zeigt einen Screenshot des Simulationsbaukastens für den Bereich der Elektrizitätslehre, der im Projekt eXkomp entwickelt wurde und eingesetzt wird. Ziel dieser Entwicklung war die Herstellung von einem Simulationsbaukasten, der in seinen Handlungsmöglichkeiten und kognitiven Anforderungen dem Realexperiment möglichst nahe kommt. Es wurde deshalb bewusst nicht das gesamte Potenzial des Mediums ausgenutzt wie es für einen Einsatz in Lehr- Lernsituationen angezeigt wäre (z.B. Einsatz von Multikodierung der Schaltkreise).

Dennoch wurden einige grundlegende Gestaltungskriterien (vgl. z.B. [24]) berücksichtigt.

Die Ablaufsteuerung der Simulation ist individuell, sodass die Informationsdichte und die kognitive Belastung individuell steuerbar ist (vgl. [24]). Um die kognitive Belastung zum Realexperiment möglichst gleich zu halten, ist das Experimentiermaterial des Simulationsbaukastens dem Material aus dem Realexperiment in Steuerung, Funktionsweise und Aussehen (bearbeitete Fotos der Experimentiermaterialien) nachempfunden. Im Regal (dunkelbrauner Bereich) liegt, wie beim Realexperiment, das Experimentiermaterial bereit. Mit dem ausgewählten Material kann dann auf der weißen Arbeitsfläche experimentiert werden. Dabei muss z.B. die Spannungsquelle eingeschaltet und die gewünschte Spannung eingestellt werden. Das Einstellen und Ablesen von Messinstrumenten ist mit dem Simulationsbaukasten schwierig. Um diesen Nachteil aufzuheben, wird z.B. die Spannungsanzeige bei Bedarf vergrößert dargestellt, um ein deutliches Ablesen zu ermöglichen (vgl. Abb. 2).



Abb.2: Simulationsbaukasten zur Elektrizitätslehre

Auch das Multimeter kann mit Kabeln versehen und der Messbereich eingestellt werden. Dabei wird der entsprechende Bereich am Multimeter vergrößert und die Messbereiche mit unterschiedlichen Farben unterlegt, um eine gute Ablesbarkeit zu ermöglichen. Es werden keine weiteren visuellen Anzeigen eingebaut, um die Probanden nicht zusätzlich kognitiv zu belasten. Das Verlegen von Kabeln wurde in verschiedenen Versionen getestet und den Wünschen der Probanden angepasst. Dabei werden Kabel verlegt, wenn der Button „Kabel legen“ angeklickt wurde und die Maus vom gewünschten Startpunkt bei gedrückter Maustaste zum gewünschten Zielpunkt bewegt wird. Während der Bewegung wird das Kabel entlang der Mausspur verlegt (gezeichnet). Der Verlauf des Kabels auf der Arbeitsfläche kann anschließend (bei festen Anschlüssen) noch mit der Maus verändert werden. Geräte können auch durch falsche Handhabung ausfallen. Sie werden durch neue Geräte im Regal ersetzt, wenn ein defek-

tes Gerät erkannt und in den Mülleimer (unten links) verschoben wurde.

Um technische Schwierigkeiten und Bedienungsprobleme zu minimieren, die zusätzlich eine kognitive Belastung darstellen (vgl. [24]), wurde der Simulationsbaukasten eingehend getestet. Bevor Probanden mit dem Simulationsbaukasten einen Test durchführen, werden sie in einer Trainingsphase mit der Simulation vertraut gemacht.

Im Hinblick auf den Einsatz der Multimediainhalte in Lernsituationen wären Optimierungen möglich. So zeigen z.B. Kalyuga et al. [25], dass verbale Zusatzinformationen die kognitive Belastung weiter verringern würde, genauso wie farbliche Kodierungshilfen. Sie stellen fest, dass es von Nachteil ist, wenn zwischen Bildbetrachtung, hier der Simulation, und Texterfassung, hier der Protokollbogen, gewechselt werden muss. Dieser Nachteil besteht allerdings auch bei den Tests mit Realexperimenten, da hier zwischen dem tatsächlichen Experiment und dem Protokollbogen gewechselt werden muss. Im Hinblick auf den geplanten Einsatz in der Testsituation wurde der Simulationsbaukasten daher diesbezüglich nicht optimiert.

3.3. Austauschbarkeit der Testformate

Shavelson et al. [26] reanalysieren empirische Daten von Baxter & Shavelson [27] mit dem Ziel, verschiedene Formen von Experimentaltests, darunter auch eine Version mit einem Simulationsbaukasten, und schriftliche Wissenstests hinsichtlich ihrer „Austauschbarkeit“ zur Erhebung experimenteller Fähigkeiten zu überprüfen. Dazu verglichen sie die erzielten Kompetenzzuweisungen aus Experimentaltests mit einem Simulationsbaukasten mit denen aus schriftlichen Wissenstests und Experimentaltests mit Realexperimenten, welche anhand von Stichwortprotokollen der Probanden oder anhand direkter Beobachtung durch trainierte Rater bewertet worden sind. Das Kategoriensystem für die Experimentaltests ist, ähnlich wie bei den TIMSS Experimentiertests (vgl. [18]), produktbezogen und sehr grobstufig. Beispielsweise wird die Bearbeitung sogenannter „Black-Box-Aufgaben“ zur Elektrizitätslehre nur dann als richtig gewertet, wenn der Inhalt richtig erkannt und die richtige Beschaltung für die Ermittlung des Inhaltes verwendet wurde. Teilerfolge oder Schritte, die zur richtigen oder falschen Lösung führten, werden nicht bewertet. Die Bewertung anhand der Beobachtung des Realexperiments wird als Maßstab für die anderen Verfahren angesetzt. Sie weist, gemittelt über verschiedene Aufgaben, die höchste Korrelation mit der Bewertung von Realexperimenten anhand der Stichwortprotokolle ($r = 0,78$) auf, die geringste Korrelation mit den schriftlichen Wissenstests ($r = 0,46$) und ebenfalls eine unerwartet geringe Korrelation mit den Experimentaltests basierend auf Computersimulation ($r = 0,50$). Die geringe Korrelation mit den schriftlichen Wissenstests führen die Autoren darauf zurück, dass die

Schülerinnen und Schüler bei den Experimentaltests eine Rückmeldung aus dem Realexperiment bzw. der Computersimulation erhalten, die in schriftlichen Wissenstests fehlt. Bei den Experimentaltests von Shavelson et al. ist zu betonen, dass die Auswertung bei allen drei Varianten produktorientiert erfolgte. Als Fazit ihrer Studien halten sie fest: “(a) *direct observation, notebook, and computer simulation are equally exchangeable, (b) paper-and-pencil methods are not exchangeable for performance assessments, and (c) student performance over occasions is volatile.*” ([26], S. 70). Teil (c) bezieht sich darauf, dass die Korrelation zwischen den Bewertungen anhand direkter Beobachtung bei Realexperimenten und Protokollen von Realexperimenten basierend auf Testdurchführungen auf 0.52 sank, nachdem zu jeweils unterschiedlichen Zeitpunkten über mehrere Monate hinweg bei einer zufällig ausgewählten Testgruppe die Tests durchgeführt wurden. Damit konnte der Einfluss des Messzeitpunktes nachgewiesen werden.

3.4. Testformate für die Vergleichsstudie eXkomp

Vor dem Hintergrund des oben beschriebenen Modells experimenteller Kompetenz sollen in der Vergleichsstudie eXkomp drei verschiedene Testformate auf ihre Eignung zur Diagnose experimenteller Kompetenz überprüft werden: Experimentaltest mit Realexperiment, schriftlicher Test sowie Experimentaltest mit Simulationsbaukasten.

3.4.1. Experimentaltest mit Realexperiment

Bei diesem Testformat erhalten die Probanden eine experimentell lösbare Aufgabenstellung (z.B. die Erfassung der Kennlinie einer Glühlampe), zugehörige grundlegende Fachinformationen (Definition der Kennlinie), einen Pool an Experimentiermaterial sowie Protokollbögen.

Die Datenerfassung erfolgt über Videoaufzeichnungen der experimentellen Handlungen sowie über die Protokollbögen. Dieses Testformat wird, wie bei Shavelson et al., als „Referenz“ für die Kompetenzzuweisung angesetzt, weil es dem realen Experimentieren (bei vorgegebener Fragestellung) sehr nahe kommt.

3.4.2. Schriftlicher Test

Hier wird die gleiche experimentelle Fragestellung mit den gleichen Fachinformationen vorgegeben. Die Probanden müssen sich in zwei fiktive Schüler, Alina und Bodo, die diese Fragestellung experimentell bearbeiten sollen, hineinversetzen. Die Schritte sind durch Multiple-Choice-Fragen, z.B. die Auswahl der richtigen Schaltung aus vier vorgegebenen Schaltungen, und einzelne offene Fragen stark vorkonstruiert. Durch einen möglichst prozeduralen Charakter des schriftlichen Tests wird versucht, eine möglichst große Validität zu erreichen.

Der Test wird online am PC bearbeitet. Die Eingaben werden bis auf eine von Hand anzufertigende Versuchsskizze direkt in einer Datenbank erfasst.

Diese Variante bietet den Vorteil, dass an bestimmten Stellen die korrekten Zwischenergebnisse, wie z.B. die richtige Schaltung, eingeblendet werden können, damit die Probanden auf dieser Basis weiterarbeiten können.

3.4.3. Experimentaltest mit Simulationsbaukasten

Anders als Shavelson [25] werden wir die Daten des Experimentaltests mit Simulationsbaukasten prozessbezogen auswerten. Analog zum ersten Testformat erhalten die Probanden die gleiche experimentell lösbare Aufgabenstellung, die Fachinformationen sowie den Protokollbogen. Anstelle des realen Experimentiermaterials wird nun ein Simulationsbaukasten (vgl. Abb. 2) eingesetzt, in dem die gleichen Geräte wie im Realexperiment und weitgehend die gleichen Handlungsmöglichkeiten zur Verfügung stehen.

Die Daten werden zunächst über eine Bildschirmaufzeichnung (Registrierung mit Software Hypercam²) sowie die Protokollbögen erfasst. Im nächsten Schritt werden von dem Simulationsbaukasten Navigationsdaten erzeugt, die eine Rekonstruktion der für die Kompetenzdiagnostik relevanten Handlungen ohne Bildschirmaufzeichnung ermöglichen soll (siehe unten).

Sowohl die Bildschirmmitschnitte als auch die Mitschnitte der Handlungen im Realexperiment werden mit Hilfe von Videograph [28] kategoriengeleitet analysiert.

Die Tabelle in Abbildung 3 stellt die drei Testformate, deren charakteristische Merkmale sowie die Erhebungsmethoden gegenüber.

Realexperiment	schriftlich	Simulationsbaukasten
• echtes Experimentieren	• in experimentelle Situation hineinversetzen	• Experimentieren innerhalb der Simulation
• Handlungsmöglichkeiten	• rein gedankliches Experimentieren	• Handlungsmöglichkeiten
• Rückkopplung	• Abwägen zwischen Distraktoren	• Rückkopplung
• Lösungsweg / Antwortformat: offen	• Lösungsweg / Antwortformat: geschlossen	• Lösungsweg / Antwortformat: offen
Erhebungsmethoden: Video & Protokoll	Erhebungsmethoden: html basiert & Papier	Erhebungsmethoden: Bildschirmaufzeichnung & Protokoll, Logdaten

Abb. 3: Gegenüberstellung der Testformate

Die Testformate sollen an drei Kriterien gemessen werden (vgl. Abschnitt 1): Sie sollen *prozessbezogene* Aspekte erfassen, einen geringen logistischen Aufwand erfordern (Versuchsmaterialien) und möglichst einfach auswertbar sein.

Beim Realexperiment ist eine prozessbezogene Auswertung möglich, Material- und Analyseaufwand sind für den großflächigen Einsatz aber groß.

² Wir verwenden eine Testversion. Diese steht zum kostenlosen Download auf <http://hypercam.softonic.de/> [25.08.2009] bereit.

Diesen Nachteil hat der schriftliche Test nicht, allerdings ist hier keine handlungsorientierte Auswertung möglich. Eine Lösung könnte der Simulationskasten sein. Wenn hier eine teilautomatisierte Auswertung mit Hilfe von Navigationsdaten [29] möglich wird, ist eine prozessbezogene Auswertung mit verhältnismäßig geringem Material- und Analyseaufwand möglich.

4. Hypothesen und Forschungsfrage

Die Hypothesen und Forschungsfragen beziehen sich auf den Vergleich von Kompetenzzuweisungen. Kompetenz beim Experimentieren messen wir an den folgenden Kriterien: Dauer bis zum Erreichen einer (Teil-) Lösung, fachliche Korrektheit der Handlungsentscheidungen und Konsistenz des Handlungsverlaufes. Diese Kriterien werden angewandt auf einzelne Merkmale experimenteller Kompetenz (vgl. Abb. 1) sowie deren Zusammenspiel. Folgefehler wirken sich deshalb nicht notwendigerweise negativ auf die Bewertung des folgenden Teilschrittes aus.

Schriftliche Wissenstests sind zwar eine sehr effiziente Erhebungsmethode, nach Shavelson et al. [26] jedoch aufgrund der fehlenden Rückkopplung aus Experiment bzw. Computersimulation nicht austauschbar gegen produktbezogen ausgewertete Experimentaltests. Daraus ergibt sich die erste Hypothese:

- a) Zwischen Kompetenzzuweisungen auf Basis von schriftlichen Wissenstests und prozessbezogenen Kompetenzzuweisungen auf Basis von Realexperimenten besteht kein hinreichend großer Zusammenhang.

Unter einem hinreichend großen Zusammenhang verstehen wir eine Korrelation von mindestens 0.7.

Shavelson et al. [26] erklären produktbezogen ausgewertete Experimentaltests mit Realexperiment bzw. Computersimulation für austauschbar. Experimentaltests mit virtuellen Medien werden bei uns gemäß der Konstruktion der Kompetenztests den Handlungsmöglichkeiten mit den Realexperimenten weitgehend entsprechen. Bei beiden Varianten können daher sehr ähnliche Handlungen durchgeführt und von Ratern basierend auf detaillierten Aufzeichnungen analysiert werden.

- b) Zwischen Kompetenzzuweisungen auf Basis der Experimentaltests mit Simulationsbaukästen und Zuweisungen, die auf Experimentaltests mit Realexperimenten beruhen, besteht ein hinreichend großer Zusammenhang.

Die Studien von Shavelson et al. zeigen zwar eine mittlere Korrelation zwischen der Bewertung von Experimentaltest anhand von Stichwortprotokollen bzw. direkter Beobachtung, jedoch sind die Beobachtungskriterien hier produktbezogen und nicht prozessbezogen. Es ist zu vermuten, dass produktbezogen ausgewertete Experimentaltests andere Aspekte experimenteller Kompetenz erfassen als pro-

zessbezogen ausgewertete. Die Qualitäten der Wege, auf denen ein bestimmtes experimentelles Ergebnis gefunden wird, können sich deutlich unterscheiden. Die Analyse der Handlungsaufzeichnungen aus Experimentaltests sollte demnach andere Ergebnisse erbringen können als die Analyse der Stichwortprotokolle. Das führt zu folgender Hypothese:

- c) Bei Realexperiment und Simulationsbaukasten besteht zwischen Kompetenzzuweisungen auf Basis von Handlungsaufzeichnungen im Vergleich zu einer Zuweisung allein auf Basis der Stichwortprotokolle kein hinreichend großer Zusammenhang.

Zusätzlich wird eine Forschungsfrage formuliert, die notwendig ist, um die Validität der Navigationsdaten im Hinblick auf teilautomatisierte Auswertungen zu überprüfen. Sie lautet:

- Wie valide erfasst bei Experimentaltests mit Simulationsbaukästen die Auswertung detaillierter Navigationsdaten die gleichen Konstrukte wie die Auswertung der Bildschirmaufzeichnungen?

5. Untersuchungsdesign

5.1. Die Kompetenztests

Die Hauptstudie (Frühjahr 2010) dient dem Vergleich der drei Testformate der Kompetenztests. Es gibt zwei Aufgabenstellungen, die sich auf die Unterrichtsgegenstände der Elektrizitätslehre der Mittelstufe beziehen. Jede Aufgabenstellung erfordert eine Abfolge unterschiedlicher experimenteller Handlungen, von der Klärung der Fragestellung hin zur Interpretation des Ergebnisses. Die Aufgabenstellungen wurden jeweils in drei Testformate (vgl. 3.3; schriftlich (P), Realexperiment (R), Simulationsbaukasten (S)) umgesetzt. Eine Aufgabenstellung besteht darin, die Kennlinie einer Glühlampe aufzunehmen. In Abbildung 4 ist die Aufgabenstellung abgebildet.

Experiment: Kennlinie einer Glühlampe

Information:

Die Eigenschaften einer Glühlampe werden durch eine Kennlinie beschrieben. Das ist ein Graph in einem Diagramm, bei dem auf der x-Achse die Stromstärke in der Glühlampe und auf der y-Achse die Spannung an der Glühlampe aufgetragen ist.

Deine Aufgabe:

Messe und zeichne die Kennlinie einer Glühlampe.

Abb. 4: Aufgabenbeispiel „Kennlinie einer Glühlampe“

Das notwendige fachliche Wissen wird bereits zu Beginn der Aufgabe bereitgestellt. Die Probanden klären zunächst die Fragestellung, entwickeln Ideen zur Durchführung, welche bei den experimentellen Varianten im Protokollbogen notiert werden sollen, und führen das Experiment durch. Ein richtiger Versuchsaufbau im Realexperiment ist in Abbildung 5, ein richtiger Versuchsaufbau im Simulationsbaukasten in Abbildung 6 zu sehen. Es wird deutlich, dass sich die Aufbauten sehr ähnlich sind.

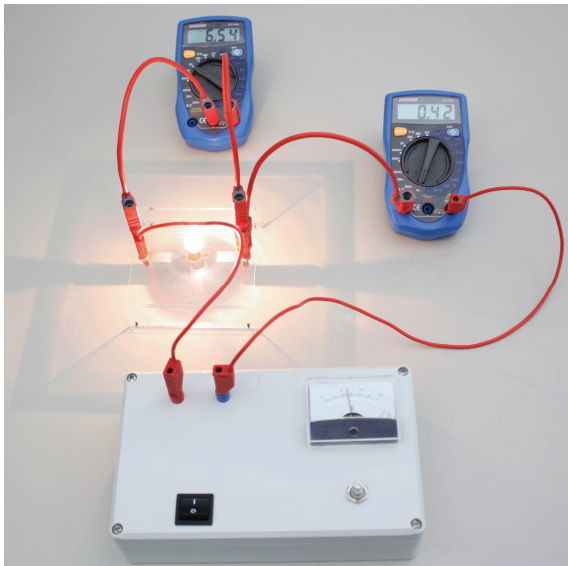


Abb. 5: Mögliche Schaltung zur Bestimmung der Kennlinie im Realexperiment

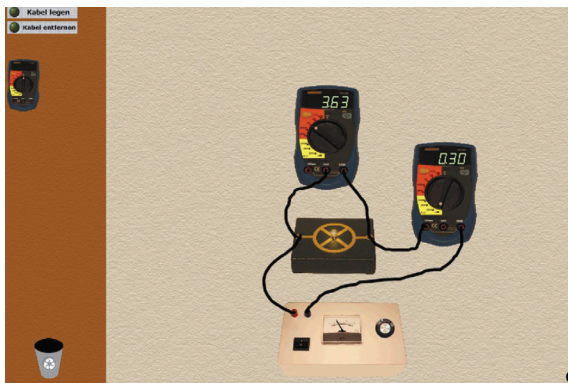


Abb. 6: Mögliche Schaltung zur Bestimmung der Kennlinie mit Simulationsbaukasten

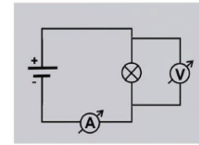
Der Protokollbogen ist weiterhin grob vorstrukturiert, indem Messwerte, Berechnungen, Ergebnisse und Begründungen verlangt werden. Bei der rein schriftlichen Variante ist die Aufgabenstellung analog, aber noch stärker vorstrukturiert. Anstelle des Schaltungsaufbaus erfolgt hier beispielsweise die Abfrage der richtigen Schaltung und der geeigneten Geräte (vgl. Abb. 7).

5.2. Begleiterhebung und Trainingsphase

Vor der Durchführung der Kompetenztests ist eine Begleiterhebung notwendig, um Personenparameter zu erfassen, von denen ein Zusammenhang mit experimenteller Kompetenz erwartet wird oder bekannt ist. Das sind kognitive Fähigkeiten [30], inhaltsbezogenes Fachwissen [31] sowie Aspekte des physikbezogenen Selbstkonzeptes, die sich speziell auf die Einschätzung der eigenen experimentellen Fähigkeiten beziehen. Außerdem werden Computerkenntnisse erfragt (Selbsteinschätzung). Zur Erfassung der kognitiven Fähigkeiten werden ausgewählte Subskalen des Kognitive-Fähigkeiten-Tests (KFT, [30]) eingesetzt.

Alina und Bodo benötigen nun Experimentiermaterialien. Kreuze an, welche sie wählen sollten.

Zur Erinnerung siehst du hier noch einmal die richtige Schaltskizze:



Alina und Bodo benötigen zusätzlich zu der zu untersuchenden Glühlampe:

	Ja	Nein
ein Netzgerät mit konstanter Stromstärke	<input type="checkbox"/>	<input type="checkbox"/>
eine Glühlampe mit Fassung auf Sockel	<input type="checkbox"/>	<input type="checkbox"/>
eine 4,5V Flachbatterie	<input type="checkbox"/>	<input type="checkbox"/>
ein regelbares Netzgerät mit Spannungsanzeige	<input type="checkbox"/>	<input type="checkbox"/>
Kabel mit Steckern	<input type="checkbox"/>	<input type="checkbox"/>

	Anzahl: 0	1	2	3
Strommessgerät	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Spannungsmessgerät	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Isolatorfuß	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Weiter

Abb. 7: Abfrage der zu wählenden Geräte beim schriftlichen Test

Da bei den Kompetenztests mit dem Simulationsbaukasten pseudo-dreidimensionale experimentelle Anordnungen in zweidimensionaler Repräsentation am PC zu bearbeiten sind, wird insbesondere das räumliche Vorstellungsvermögen berücksichtigt. Darüber hinaus werden mit einer weiteren Subskala die sprachlichen Fähigkeiten als Voraussetzung für das Verständnis der Aufgabenstellungen erfasst. Für den modifizierten Fachwissenstest werden Tests von von Rhöneck [32] und Beichner & Engelhardt [33] zugrunde gelegt, die speziell auf bekannte Lernschwierigkeiten im Bereich „elektrischer Stromkreis“ eingehen. Das Selbstkonzept bezüglich experimenteller Fähigkeiten und die Selbsteinschätzung werden in einem modifizierten Fragebogen nach Brell [34] erhoben. Die Daten dienen der Zuweisung der Probanden in vergleichbare Gruppen für die Kompetenztests (matched samples).

Neben der Begleiterhebung ist eine Trainingsphase notwendig, in der alle Probanden sowohl mit dem Simulationsbaukasten als auch mit den Experimentiermaterialien vertraut gemacht werden. Da die Trainingsphase zwei Wochen vor der Haupterhebung stattfindet, ergibt sich hieraus eine größere Flexibilität für die Gruppeneinteilung. Es wird eine für das Training entwickelte zusätzliche Aufgabenstellung zum Thema Elektrizität mit dem Simulationsbaukasten und dem Realexperiment bearbeitet.

5.3. Pilotstudien

Eine erste Pilotstudie diente der Optimierung der Übungsmaterialien, der Begleiterhebung, der Kompetenztests und der Erprobung von Auswertungsmethoden. Für die Datenerhebung (März/April 2009) waren ursprünglich 30 Probanden der Jahrgangsstufe 10 (Gymnasium) eingeplant. Die Aufgabenstellungen aus dem Themenbereich „elektrische Stromkreise“ erwiesen sich als unerwartet schwierig. Den

Probanden der 10. Jahrgangsstufe fehlen zentrale Konzepte der Elektrizitätslehre, z.B. das Konzept des elektrischen Widerstands. Hinzu kamen grundlegende Probleme beim Umgang mit Multimetern und Spannungsquellen. Es wurden daher zusätzlich Schülerinnen und Schüler der 12. Jahrgangsstufe einbezogen, die diese Probleme teilweise auch hatten, jedoch in abgeschwächter Form.

Die zweite Pilotstudie dient der weiteren Erprobung und Optimierung der Kompetenztests sowie der Erprobung und Validierung der Verfahren zur Aufzeichnung und zur Auswertung der Navigationsdaten, die bei der Computersimulation erzeugt werden.

5.4. Hauptstudie

Das Design der Hauptstudie ist in Abbildung 8 schematisch dargestellt.

Schule	Begleiterhebung			
Didaktiklabor	Trainingsphase			
	Tests experimenteller Kompetenz			
Uni Bremen	Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
	R1	S1	R1	P1
	S2	R2	P2	R2

R: Realexperiment, P: schriftlich, S: Simulation

1: Aufgabenstellung 1, 2: Aufgabenstellung 2

Abb. 8: Design der Hauptstudie

Die Hauptstudie gliedert sich in drei Phasen: die Begleiterhebung an der Schule und die Trainingsphase sowie die Durchführung der Kompetenztests im Physikdidaktiklabor der Universität Bremen. Die Begleiterhebung (siehe Abschnitt 5.2) wird im Rahmen einer Schulstunde in Klassen- bzw. Kursverbänden mit insgesamt ca. 180 Probanden durchgeführt. Der Überhang von 50% gegenüber der Probandenzahl der Kompetenztests ermöglicht es, Probanden mit extremen Personenparametern bei der Zusammenstellung der Testgruppen auszuschließen und Ausfälle, z.B. durch Krankheit oder Klassenwechsel, zu kompensieren.

Die Trainingsphase zur Einführung der Medien und Arbeitsformen wird ca. zwei Wochen vor der Durchführung der Kompetenztests im Didaktiklabor durchgeführt.

Anhand der erhobenen Personenparameter werden für die Kompetenztests vier Gruppen mit je ca. 30 Probanden zusammengestellt (matched samples). Die Einteilung erfolgt auf Basis der Individuen. Gleichzeitig ist sicherzustellen, dass jede Klasse auf alle vier Gruppen aufgeteilt wird.

Die Kompetenztests werden mit beiden zuvor erprobten Aufgabenstellungen (siehe Abschnitt 5.1) durchgeführt. Sie finden im Didaktiklabor statt, wo die für die Experimentaltests notwendigen apparativen Voraussetzungen in hinreichender Anzahl zur

Verfügung stehen (Experimentiermaterial, PC-Arbeitsplätze, Kameras). Um den Einfluss der Umgebung für alle Probanden gleich zu halten, finden auch die schriftlichen Wissenstests dort statt.

Da der Experimentaltest mit Realexperiment (R) als Referenz für die Qualität der Testformate gewählt wurde, bearbeitet jeder Proband eine der beiden Aufgabenstellungen in diesem Testformat (R) und die zweite Aufgabenstellung in einem der anderen beiden Testformate (S oder P). Aufgrund des von Shavelson et al. nachgewiesenen Einflusses des Messzeitpunktes wird zusätzlich ein Kreuzdesign gewählt (vgl. Abb. 8).

5.5. Teilstudien zur Auswertung

Zur Überprüfung der Hypothesen und Beantwortung der Forschungsfrage werden im Rahmen der Pilotstudien und der Hauptstudie entsprechende Teilstudien durchgeführt.

- Zur Überprüfung der ersten Hypothese wird für alle Probanden der Gruppen drei und vier anhand der schriftlichen Wissenstests zum Experimentieren eine Kompetenzzuweisung getroffen. Eine weitere Kompetenzzuweisung wird anhand der Analysen der Handlungsaufzeichnungen beim Realexperiment getroffen. Die damit erzielten Kompetenzzuweisungen werden auf Zusammenhänge untersucht.
- Zur Überprüfung der zweiten Hypothese werden für alle Probanden der Gruppen eins und zwei anhand der Analysen der Handlungsaufzeichnungen für jede Testvariante des Realexperiments und des Simulationsbaukastens Kompetenzzuweisungen getroffen. Diese werden auf Zusammenhänge untersucht.
- Zur Überprüfung der dritten Hypothese wird zunächst für alle Probanden für alle Testdurchführungen mit Realexperiment oder Simulationsbaukasten anhand der Stichwortprotokolle eine Kompetenzzuweisung getroffen. Basierend auf der Handlungsaufzeichnung wird eine weitere Kompetenzzuweisung getroffen. Beide Zuweisungen werden auf Zusammenhänge untersucht.
- Zur Beantwortung der Forschungsfrage werden für alle Probanden, die Kompetenztests mit dem Simulationsbaukasten bearbeitet haben, zum einen anhand der Bildschirmaufzeichnungen und zum anderen anhand der Navigationsdaten Kompetenzzuweisungen getroffen. Die mit den verschiedenen Verfahren erzielten Kompetenzzuweisungen werden auf Zusammenhänge untersucht.

6. Ausblick

Wenn die untersuchten technologiegestützten Verfahren sich als valide erweisen, wird es u.a. möglich, in Studien zur Modellierung naturwissenschaftlicher Kompetenz auf breiter empirischer Datenbasis das Zusammenspiel der experimentellen Kompetenz mit anderen Komponenten naturwissenschaftlicher Kompetenz zu untersuchen. Für die Auswertung müssten dann zusätzlich Algorithmen zur Analyse von Log-Daten entwickelt werden.

Auch das Spektrum der Prüfungsformen, die in der Schulpraxis für experimentelle Kompetenzen zur Verfügung stehen, würde erweitert.

7. Literatur

- [1] KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.) (2004). Bildungsstandards im Fach Physik (Chemie & Biologie) für den Mittleren Schulabschluss. München: Luchterhand.
- [2] Schecker, H., Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. Zeitschrift für Didaktik der Naturwissenschaften 12, 45-66.
- [3] Neumann, K., Kauertz, A., Lau, A., Notarp, H. & Fischer, H. E. (2007). Die Modellierung physikalischer Kompetenz und ihrer Entwicklung. Zeitschrift für Didaktik der Naturwissenschaften, 13, 125-143.
- [4] Labudde, P. & Stebler, R. (1999): Lern- und Prüfungsaufgaben für den Physikunterricht. Erträge aus dem TIMSS-Experimentiertest. In: Unterricht Physik 10 Nr. 54, 23-31.
- [5] Harmon, M., Smith, T. A., Martin, M. O., Kelly, D. L., Beaton, A. E., Mullis, I. V. et al. (1997). Performance Assessment in IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: TIMSS International Study Center, Boston College.
- [6] Adamina, M., Labudde, P., Gingins, F., Nidegger, C., Bazzigher, L., Bringold, B. et al. (2008). HarmoS Naturwissenschaften: Kompetenzmodell und Vorschläge für Bildungsstandards (Wissenschaftlicher Schlussbericht). Bern.
- [7] Goldberg, F., Bendall, S. (1995). Making the invisible visible: A teaching and learning environment that builds on a new view of the physics learner. American Journal of Physics 63, 11, 978-991.
- [8] Höttecke, D. (2007). Fachliche Klärung des Experimentierens. In Höttecke, D. (Hrsg.), Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung. Münster: LIT, 293-295.
- [9] Hammann, M. (2004). Kompetenzentwicklungsmodelle: Merkmale und ihre Bedeutung – dargestellt anhand von Kompetenzen beim Experimentieren. Der mathematische und naturwissenschaftliche Unterricht 57, Nr. 4, 196-203.
- [10] Klahr, D. (2000). Exploring Science. The cognition and development of discovery processes. Massachusetts.
- [11] Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger, H. Vogt (Hrsg.), Theorien in der biologiedidaktischen Forschung, Berlin: Springer.
- [12] Möller, A., Grube, C. & Mayer J. (2007). Kompetenzniveaus der Erkenntnisgewinnung bei Schülerinnen und Schülern der Sekundarstufe I. In Bayrhuber et al. (Hrsg.), Ausbildung und Professionalisierung von Lehrkräften, Internationale Tagung der Fachgruppe Biologiedidaktik im Verband Biologie, Biowissenschaften & Biomedizin; Essen.
- [13] Dickmann, M. (2009). Validierung eines computergestützten Experimentaltests zur Diagnostik experimenteller Kompetenz. Nicht veröffentlichte Bachelorarbeit, Technische Universität Dortmund.
- [14] Holländer, L. K. (2009). Validierung eines Experimentaltests mit Realexperimenten zur Diagnostik experimenteller Kompetenz. Nicht veröffentlichte Bachelorarbeit, Technische Universität Dortmund.
- [15] Walpuski, M. (2006). Optimierung von Kleingruppenarbeit durch Strukturierungshilfen und Feedback. Band 49. In Niedderer, H.; Fischler, H.; Sumfleth E. (Hrsg), Studien zum Physik- und Chemielernen. Berlin: Logos.
- [16] Neumann, K. (2004). Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker. In Niedderer, H.; Fischler, H. (Hrsg.), Studien zum Physiklernen, Bd. 38. Berlin: Logos.
- [17] Aufschnaiter, C. von (2002). Ich weiß was rauskommt, aber ich kann es nicht erklären. In Pitton, A. (Hrsg.), Außerschulisches Lernen in Physik und Chemie (S. 114–116). Münster: LIT.
- [18] Stebler, R., Reusser, K., Ramseier, E. (1998). Praktische Anwendungsaufgaben zur integrierten Förderung formaler und materialer Kompetenzen – Erträge aus dem TIMSS Experimentiertest. Bildungsforschung und Bildungspraxis, 20(1), 28-54.
- [19] Berge, O. E., Volkmer, M. (2002). Lernerfolgskontrolle mit Experimenten – Didaktische Aspekte. Naturwissenschaften im Unterricht – Physik, 71/72/2002, 206-210.
- [20] Henke, C. und Sumfleth, E. (2005). Leistungsmessung in der Oberstufe mit chemischen Experimentalaufgaben. In Pitton, A. (Hrsg.), Lehren und Lernen mit neuen Medien. Münster: LIT, 340-342.
- [21] Klos, S. & Sumfleth, E. (2006). Einfluss des Fachs Naturwissenschaft auf Lernerfolg und Interesse. In Pitton, A. (Hrsg.), Lehren und Lernen mit neuen Medien. Gesellschaft für Didak-

- tik der Chemie und Physik. Münster: LIT, 343-345.
- [22] Ruman, S. (2005). Kooperatives Experimentieren im Chemieunterricht. Entwicklung und Evaluation einer Interventionsstudie zur Säure-Base-Thematik. Berlin: Logos.
- [23] Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrum, R., Rolff, H.-G., Rost, J., Schiefele, U. (2004). PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. Münster: Waxmann.
- [24] Girwidz, R. (2004). Lerntheoretische Konzepte für Multimediaanwendungen zur Physik. Physik und Didaktik in Schule und Hochschule 1/3, 9-19.
- [25] Kalyuga, S., Chandler, P., Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied-Cognitive-Psychology* 13(4), 351-371.
- [26] Shavelson, R. J., Ruiz-Primo, M. A., Wiley, E. W. (1999). Note on Sources of Sampling Variability in Science Performance Assessments. *Journal of Educational Measurement*, 36 (1) 61-71.
- [27] Baxter, G. P., Shavelson, R. J. (1994). Science Performance Assessments: Benchmarks and Surrogates. *International Journal of Educational Research*, 21, 233-350.
- [28] Rimmele, R. (2008). Videograph® (Version 4.1.4.1.X3 für Windows).
- [29] Zastrow, M. (2001). Interaktive Experimentieranleitungen. Entwicklung und Evaluation eines Konzeptes zur Vorbereitung auf das Experimentieren mit Messgeräten im Physikalischen Praktikum. In Niedderer, H.; Fischler, H. (Hrsg.), *Studien zum Physiklernen*, Band 18. Berlin: Logos.
- [30] Heller, K. A. & Perleth, C. (2000). Kognitiver Fähigkeitstest für 4.-12. Klassen, Revision (KFT 4-12+ R). Göttingen: Hogrefe.
- [31] Henke, C. (2007). Experimentell – naturwissenschaftliche Arbeitsweisen in der Oberstufe. In Niedderer, H.; Fischler, H. (Hrsg.), *Studien zum Physiklernen*, Bd. 59, Berlin: Logos.
- [32] Rhöneck, C. von (1988): Aufgaben zum Spannungsbegriff. *Naturwissenschaften im Unterricht - Physik/Chemie* 36 Nr. 31, S. 38-41.
- [33] Engelhardt, P.V. & Beichner, R.J. (2004): DIRECT – Determining and Interpreting Resistive Electric Circuits Concepts Test. In: *American Journal of Physics*, Vol. 72, 1, 98-115.
- [34] Brell, C. (2008). Lernmedien und Lernerfolg – reale und virtuelle Materialien im Physikunterricht. In Niedderer, H.; Fischler, H. (Hrsg.), *Studien zum Physiklernen*, Bd. 38, Berlin: Logos.